

# Demo: Real-time Remote Reporting of Active Regions with Wi-FLIP

Jorge Fernández-Berni <sup>\*</sup>, Ricardo Carmona-Galán <sup>\*</sup>, Gustavo Liñán-Cembrano <sup>\*</sup>,  
 Ákos Zarándy <sup>†</sup>, Ángel Rodríguez-Vázquez <sup>\*</sup>

<sup>\*</sup> Institute of Microelectronics of Seville (IMSE-CNM), Seville, Spain  
 Consejo Superior de Investigaciones Científicas y Universidad de Sevilla

<sup>†</sup> Computer and Automation Research Institute (MTA-SZTAKI), Budapest, Hungary

Contact email: berni@imse-cnm.csic.es

**Abstract**—This paper describes a real-time application programmed into *Wi-FLIP*, a wireless smart camera resulting from the integration of *FLIP-Q*, a focal-plane low-power image processor, and *Imote2*, a commercial WSN platform. The application, though simple, shows the potentiality of the reduced scene representations achievable at *FLIP-Q* to speed up the processing. It consists of detecting the active regions within the scene being surveyed, that is, those regions undergoing thresholded variations with respect to the background. If an activity pattern is prescribed, *FLIP-Q* enables the reconfigurability of the image plane accordingly, making its detection and tracking easier. For each frame, the number of active regions is calculated and wirelessly reported in real time. A base station picks up the radio signal and sends the information to a PC via USB, also in real time. Frame rates up to around 10fps have been achieved, although it greatly depends on the light conditions and the image plane division grid.

## I. INTRODUCTION

Image simplification is a key issue when it comes to implement a certain algorithm on vision-enabled WSN nodes [1]–[3]. It permits to alleviate the resource allocation during the processing and hence to reduce the power consumption. According to the computer vision framework [4], early vision represents the processing stage where the greatest computational effort must be done in order to attain an adequate scene representation for further understanding in subsequent stages. Alternative architectures can be proposed to handle early vision more efficiently than conventional imager-memory-DSP architectures. For example, by making the most of the ability of CMOS processes to integrate pure imaging with signal processing circuitry. Thus, the imager becomes not only an array of photosensitive devices, but an actual processing lattice composed of elementary processing units working concurrently with photosensing [5]–[7]. This is the approach supporting the design and implementation of *FLIP-Q*, a prototype chip designed ad-hoc for ultra low-power applications [8]. This chip consists of a SIMD-based focal-plane array implementing a subset of early vision processing primitives intended to deliver a very reduced data flow. It makes the computational load of the subsequent digital processor much lighter, reducing significantly its clock frequency and the number of memory accesses and, consequently, its power consumption. In our case, the digital processor is that of *Imote2* (MEMSIC Inc.), the commercial WSN platform which has been integrated

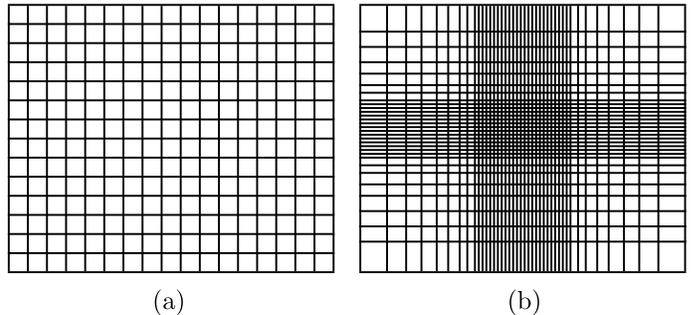


Fig. 1. Regular (a) and content-aware (b) image plane division.

with *FLIP-Q* in order to obtain *Wi-FLIP*, a low-power vision-enabled mote [9].

## II. RECONFIGURABILITY OF THE IMAGE PLANE

A way of reducing the representation of a scene is to pay careful attention only to those regions that at any moment and for any reason are of interest. The rest of the scene keeps being surveyed, but with a coarser detail. To this end, a very useful tool consists of a dynamic content-aware image plane division. Consider for example a very simple application: the detection of active regions within a scene. The objective is to provide the number of such regions,  $N$ , as an indicator of the global activity. It could be useful to initiate a thorough analysis of the scene only when the value of this indicator exceeds a certain threshold. A simple option to carry out this task is to establish a regular division of the image plane into blocks of  $B \times B$  pixels, as depicted in Fig 1(a). Each block is represented by the mean value of its pixels, which is tracked in order to detect noticeable changes. Thus, being  $I_{B_{kl}}$  the value of the block  $(k, l)$  corresponding to the background representation and  $I_{F_{kl}}$  the value of that same block but corresponding now to the foreground representation, the condition to consider activity within the block will be:

$$|I_{B_{kl}} - I_{F_{kl}}| \geq T \quad (1)$$

where  $T$  represents the global threshold determining the sensitivity to changes within the blocks.

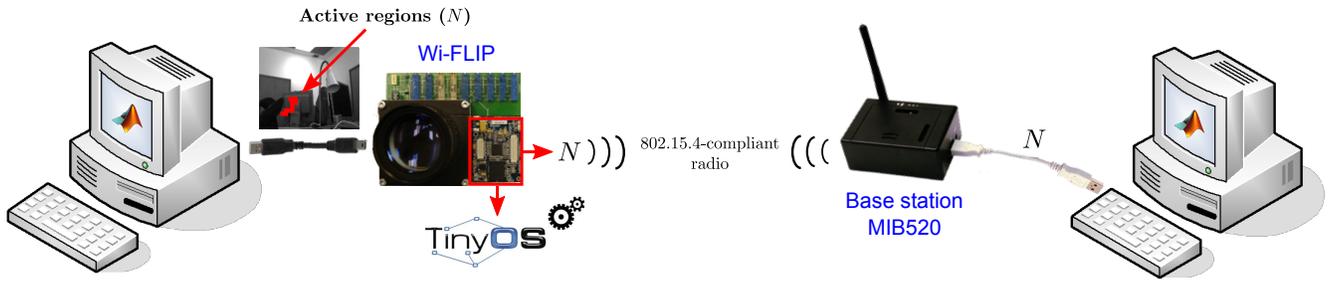


Fig. 2. General scheme of the demo.

Let us assume now that a certain pattern of activity is expected around the center of the scene. A finer sensitivity is therefore desired around that zone whereas, as getting farther, a coarser sensitivity suffices. A possibility would be to define not a global but a local threshold depending on the location of the block analyzed. However, a much simpler and efficient option is possible provided that the image plane can be reconfigured as depicted in Fig. 1(b). In that case, we could keep a global threshold since the image representation is already taking into account the targeted pattern of activity. Fewer pixels are grouped around the region of interest, allowing for finer tracking. As getting farther from that region, progressively coarser information is delivered. The point is that this reconfigurability of the image plane is provided by *FLIP-Q* at ultra low energy cost (around 20nJ estimated).

### III. DEMO SETUP

The general scheme of the demo is depicted in Fig. 2. A *nesC*-coded *TinyOS* application is running on PXA271, the 32-bit ARM5 processor of the *Imote2* platform. This application consists basically of establishing a certain configuration of the image plane at *FLIP-Q* and subsequently capturing frames. For the sake of simplicity, the first frame is considered to be the background representation of the scene. The next ones are compared to this first frame at the *Imote2*'s processor in order to detect active regions according to Eq. (1). Interestingly, the simplified images associated with the focal plane configuration previously set are available at *FLIP-Q* immediately after photointegration. That is, apart from the exposure time, no extra time is required to obtain them, speeding up significantly the processing. For each frame,  $N$  is also calculated by the ARM5 processor and included in the payload of a radio package which is broadcast. Then, a base station (MIB520 from MEMSIC Inc.) picks up that radio package and sends, through a serial port mapped into a USB connection, the information to a PC for further processing and visualization. All this operation takes place in real time. Also the background of the scene along with the location of the regions activated are sent from *Wi-FLIP* to a PC via USB. Unfortunately, this information, much heavier than the single value of  $N$  at the base station, can not be visualized in real time because of the software overhead introduced by MATLAB.

### ACKNOWLEDGMENT

This work is partially funded by the Andalusian regional government through project 2006-TIC-2352, the Spanish Ministry of Science through project TEC 2009-11812, co-funded by the European Regional Development Fund, and by the Office of Naval Research (USA), through grant N000141110312.

### REFERENCES

- [1] M. Rahimi, R. Baer, O. Iroezzi, J. Garcia, J. Warrior, D. Estrin, and M. Srivastava, "Cyclops: In situ image sensing and interpretation in wireless sensor networks," in *Proc. of 3rd Int. Conf. on Embedded Networked Sensor Systems (SenSys)*, 2005, pp. 192–204.
- [2] S. Hengstler, D. Prashanth, F. Sufen, and H. Aghajan, "Mesheye: a hybrid-resolution smart camera mote for applications in distributed intelligent surveillance," in *Proc. of 6th Int. Conf. on Information Processing in Sensor Networks*, 2007, pp. 360–369.
- [3] A. Rowe, D. Goel, and R. Rajkumar, "Firefly mosaic: A vision-enabled wireless sensor networking system," in *28th IEEE International Real-Time Systems Symposium, RTSS*, 2007, pp. 459–468.
- [4] R. González and R. Woods, *Digital Image Processing*. Prentice Hall, 2002.
- [5] G. Liñán Cembrano, A. Rodríguez-Vázquez, R. Carmona-Galán, F. Jiménez-Garrido, S. Espejo, and R. Domínguez-Castro, "A 1000 FPS at 128x128 vision processor with 8-bit digitized I/O," *IEEE J. of Solid-State Circuits*, vol. 39, no. 7, pp. 1044–1055, 2004.
- [6] P. Dudek and P. Hicks, "A general-purpose processor-per-pixel analog SIMD vision chip," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 1, pp. 13–20, 2005.
- [7] J. Poikonen, M. Laiho, and A. Paasio, "MIPA4k: A 64x64 cell mixed-mode image processor array," in *ISCAS*, 2009, pp. 1927–1930.
- [8] J. Fernández-Berni, R. Carmona-Galán, and L. Carranza-González, "FLIP-Q: A QCIF resolution focal-plane array for low-power image processing," *IEEE J. of Solid-State Circuits*, vol. 46, no. 3, pp. 669–680, 2011.
- [9] J. Fernández-Berni, R. Carmona-Galán, G. Liñán Cembrano, A. Zárandy, and A. Rodríguez-Vázquez, "Wi-FLIP: A wireless smart camera based on a focal-plane low-power image processor," in *IEEE/ACM Int. Conf. on Distributed Smart Cameras*, 2011, accepted for oral presentation.